UNIVERSIDADE NOVA DE LISBOA · NOVA SCHOOL OF BUSINESS AND ECONOMICS

Morgan Stanley

Estimating Regulatory Capital Model Parameters for Securitized Products under Basel II/III

Luís Sá Carneiro Ferreira Meirinhos · 622

May 26, 2014

Masters in Finance · Working Project
CEMS Masters in International Management · Business Project

This page is intentionally left blank

CONTENTS

1. BRIEF CONTEXT OF THE BUSINESS PROJECT 1
1.1. Company & Market Overview 1
1.2. Current Situation of Morgan Stanley 1
1.3. The Business Project Challenge & Conclusions 2
2. FURTHER DEVELOPMENT OF A SPECIFIC TOPIC 3
2.1. Original Approach 3
2.2. Main Limitations 3
2.3. Selecting Independent Variables for the PD Model 4
2.3.1. Collecting potential variables 4
2.3.2. Process to test the potential variables 6
2.3.3. Ideal number of variables 8
2.4. Conclusions 8
3. REFLECTION ON LEARNING 9
3.1. Masters Content Applied 9
3.2. New Knowledge 10
3.3. Personal Experience 10
3.4. Benefit of hindsight 10
4. REFERENCES 11

1. BRIEF CONTEXT OF THE BUSINESS PROJECT

In this section, I present the work done for the CEMS business project at Corvinus University of Budapest. To do so, I first introduce the corporate partner – Morgan Stanley (MS) – and its current situation, as well as the market in which the company operates. I then move on to the explanation of the challenge, the used methodology and a summary of the conclusions.

1.1. Company & Market Overview

Morgan Stanley is the 4th biggest investment bank in the world (The Financial Times, 2014). Headquartered in New York, the firm provides a wide set of financial services in 42 different countries around the world. One of those services consists of financing other institutions that invest large amounts of money in structured products such as Mortgage-backed securities (MBS). This is the area our business project focused on.

Such activity naturally involves a certain amount of risk; therefore risk management plays an essential role within the operations of investment banks like MS and also to the stability of the financial system, given the systemic impact of financial institutions in case of default. In fact, there is not only a entire department dedicated to risk management within each investment bank, but also several regulatory institutions such as the Basel Committee on Banking Supervision (BCBS) that dedicate their resources to stabilizing the banking system.

One of the most important initiatives of the BCBS was the creation of the Basel Accord in 1988 – essentially a document with a set of rules for the minimum capital requirements[1] of banks. This document was subsequently improved and republished under the name of Basel II. However, due to the subprime financial crisis, in 2008, the second version of the Basel Accord was criticized, resulting in an updated version of the Basel II (Basel 2.5) and the introduction of the Basel III that complements the previous one.

In this context, financial institutions are now embracing the changing environment of risk management and optimizing their operations according to the new regulation.

1.2. Current Situation of Morgan Stanley

Morgan Stanley is no exception to the current weave of adaptation to the new risk management standards. Indeed, with the recovering of the market confidence in MBS (The

[1] Banks who lend money on a risk basis are required to set apart a specific amount of capital to eliminate the risk of bankruptcy in case the borrower fails to pay back. The total amount of capital banks must hold for that purpose is referred to as the minimum capital requirements.

Economist, 2014), MS is planning to increase its exposure to securitized products. Consequently, the company is now in the process of building and approving the financial models to calculate the minimum capital requirements for each type of product. The optimization of these models is essential as holding capital that cannot be invested represents a large opportunity cost for banks.

1.3. The Business Project Challenge & Conclusions

For the business project, MS handled my team the latest model to measure the probability of default of UK MBS, still under development, and asked us to further explore and compare the logistic regression vs. the linear regression as well as to find ways to improve the overall model. For confidential reasons, MS did not share the meaning of the variables or the exact purpose of the model; hence the analysis was merely statistical.

To tackle the problem, we started with a thorough analysis of current model as well as a review of the current literature on risk management modeling. This was followed by a review of the linear and the logistic Regressions. At this stage and with the help of STATA we were able to compare both models when applied to the data provided by MS. After this comparison we explored other specific topics of the model, such as sample size and transformations, in an attempt to find possible improvements for the model.

The first important conclusion of the business project was the greater suitability of the logistic regression for this specific model. Firstly, because of the superior fit of the logistic s-shaped function when compared to the linear function. Secondly, because in this model we are analyzing a binary outcome – default or non-default – and the logistic regression was specifically designed to model binary outcomes. Thirdly, because the logistic regression estimates the coefficients by maximizing the likelihood of obtaining the observed variables, which goes in line with what we want to estimate – the likelihood of default – rather than minimizing squared sum of the errors.

Additionally, we found that inserting transformations of the independent variables on the model such as interactions has a great potential of improving its prediction power. Moreover, given the reduced size of the sample, we suggested that an increase in the size of the dataset would have a significant positive impact on the stability of the model. Finally, we concluded the project by introducing three different advanced tools that have consistently proved to deliver superior results when compared to the traditional regression methods – discriminate analysis, decision trees and neural networks. We recommended MS to further explore them.

2. FURTHER DEVELOPMENT OF A SPECIFIC TOPIC

Throughout the whole project, we took all the independent variables as given, however there is a whole process behind the selection of variables for a model, and a good selection process has indeed a significant impact on the model's quality. In this section, I further explore MS's model by studying the good practices of the independent variable selection process that could be applied in this particular case. This analysis could bring real value added for MS. I first start by describing the original approach that was used in this model to select the variables. I then present the main limitations of the approach used and do a thorough analysis of the good practices of the selection of covariates.

2.1. Original Approach

As mentioned before, the business project focused purely on statistical methods. The goal was to optimize and improve the model from a statistical point of view while considering the whole dataset as given and not taking into account the meaning of each variable. In fact, for confidential reasons the six variables were codified into V1, V2,…, V6, so we didn't even know what each variable represented.

For the very first draft of the model to predict the probability of default of the UK MBS, Morgan Stanley used some of the variables that were present in their U.S. model for other securities. Although it is a good starting point to use this simplistic approach to select the variables, there is clearly room for improvement. There are certainly a few variables missing that could improve the model, furthermore, the risk of MBS completely changes according to the reality of the respective country.

2.2. Main Limitations

The fact that we didn't know the meaning of the variables naturally constituted a real limitation to our statistical analysis. By not understanding the meaning of the variables, we could not properly assess the estimated coefficients of the different models. Moreover, it also restricted the potential for the statistical optimization of the model. For example, the addition of interactions could help us improve the model. However, each interaction provides an answer for different questions and those questions have to make economic sense in order to add value to the model, or else they will just add autocorrelation and noise (this will be further discussed below).

Also, the way Morgan Stanley selected the variables for the first model was rather basic since they did not use a structured method for selecting and testing independent variables, and only used 6 binary variables. This raises the questions: is there room for improvement, and if yes, how?

2.3. Selecting Independent Variables for the PD Model

2.3.1. Collecting potential variables

There are several techniques to model the probability of default of credit. A few examples are the logistic and the probit regressions, discriminate analysis and neural networks. Which model yields the best results can be subject of a long discussion, but there is one thing they all have in common: they are all based on historical financial data. As a result, we have to be able to decide what financial data is relevant for the model and how to select it.

The objective is to choose the variables that result in the best model possible to predict the probability of default. This selection process can be divided into two parts: firstly, one has to select the potential variables for the model; then, we must assess the different variables within the model to filter those which are justified to be present in the model. Nevertheless, we should keep in mind that the process of selecting variables for a model is not a precise science; it also involves experience and common sense (Hosmer, Lemeshow, & Sturdivant, 2013).

Before starting to collect the potential inputs, banks must take into account the Basel Accord requirements and recommendations regarding inputs collection. The accord emphasizes the importance of the quality of the data. It states that the banks must have a robust process for vetting the data inputs into the model that includes an assessment of the accuracy, completeness and appropriateness of the data. Additionally, financial institutions must demonstrate that the data represent the actual population of the actual borrowers or facilities. In the case of Morgan Stanley, they have to make sure that the input variables represent the population of the UK MBS investors or the underlying mortgages that compose the package (Basel Committee on Banking Supervision, 2006). I recommend paying special attention in this step, given the complexity behind such structured products. The Basel Accord also requires financial institutions to adopt a “Data Maintenance Process”. Banks do not only have to collect data having in mind the previous recommendations, but they also have to store them in a way that provides enough detail to allow to retrospectively reallocate obligors and facilities to grades (Basel Committee on Banking Supervision, 2006).

There are several studies of famous institutions around the world in which variables are relatively relevant for the estimation of the probability of default on credit. The TCRI states that a company is successful when it is profitable, manages its assets efficiently, has a sound financial planning and is a market leader. This suggests that to study the probability of default of institutions, one should collect variables on four dimensions: profitability, efficiency, security and size. (Fang & Huang, 2011). Another model complements these four dimensions with another two: liquidity and growth ability (Falkenstein, 2000). These variables can easily be found in the financial reports of the companies.

In the previous paragraph, I suggest a set of financial data that could be used to estimate the probability of default of an institution. Nevertheless, according to the empirical evidence on TCRI in credit rating, financial data alone fails to reflect all the credit risks of a company as many of those risks are actually mirrored in non-financial inputs. Examples of such variables are external opinions from rating agencies and auditing firms, worldwide presence. One of the main arguments in favor of the use of non-financial variables is that the financial reports may not reveal the real situation of the company as there are several ways for companies to legally manipulate and distort their books in order to make them look better (Fang & Huang, 2011).

Moreover, it is not only important to collect variables regarding the financial institutions that invest in MBS, but also regarding the people who actually ask for the mortgage – borrower specific parameters –, as well as the inputs that are particular to mortgages – mortgage specific parameters. Those inputs are the base of the pyramid that constitutes the MBS and therefore have a high potential of being important predictors of the probability of default for the UK MBS. Examples of such covariates could be: loan-to-value (LTV), debt-to-income (DTI), demographics such as age or occupation, etc. (Campbell & Cocco, 2011) (Donchev, 2009) (Neuenschwander & Proffitt, 2013).

Another important step in the process of collecting variables for the model is to include macroeconomic variables. Models that fail to include macroeconomic information run the risk of yielding imprecise PD estimations (Engelmann & Porath, 2012). Examples of macroeconomics variables are the GDP, the interest rate, the exchange rate, the house price index, the unemployment rate, etc.

At this stage, the bank should have a large set of data to start building the model. However, the quality of the data varies significantly and the way it is displayed and organized might not be the best. Before testing the variables in the model, one should observe the data, analyze and interpret it – registering the expected effect of the variable on the model. After this process, in case the analyst notices low accuracy problems, outliers in the data or other problems, the data should be treated resorting to statistical techniques. Skipping this step can significantly lower the predictability power of the model. A common technique used to model the probability of default is applying the log function to the series. This has proven to improve the results in certain situations (Blanco, Irimia, & Olive, 2012). In the model provided by Morgan Stanley, a binary categorization was applied to the variables, for example, if the unemployment rate was above a specific threshold then the variable would be 1, otherwise its value would be 0. This technique eliminates many problems such as the presence of outliers or a weak precision in measurement.

Morgan Stanley's model has 6 variables that were categorized into binary variables. Because the meaning of the variables were hidden for confidential reasons, it is impossible analyze the current selection of variables. Nonetheless, MS should make sure it includes variables of the three mentioned dimensions: company-specific, borrower-specific and macroeconomic variables. Assuming that the binary categorization of the variables was made to overcome quality problems with the data, MS could analyze if categorizing them into discrete variables of more than 2 values could add explanatory power to the model. Treatment is important, but by categorizing the variables into binary inputs, they are losing a lot of information that could potentially improve the model.

2.3.2. Process to test the potential variables

After selecting all possible variables, we need to assess them in the model in order to test which ones are significant and actually contribute to the model. The explanatory power of each variable changes according to the used statistical method. The Logistic Regression was the method used in my business project, additionally, it is considered the best model to estimate PD (Fang & Huang, 2011); therefore, in this subsection, I will present a 7-step process to test the potential variables for the logistic regression.

In a first step, we should start with a careful univariable analysis of each independent variable against the dependent variable. In the case of MS' binary input variables, this could be done using a standard contingency table analysis of the outcome versus the 2 levels of independent variables (2x2 table). Using the results of these univariable regressions we can identify the candidates for a first multivariate model by selecting all the variables with a p-value less than 0.25, along with all the variables that must be present in the model due to their clinical importance.

In the 2nd step, we finally run the multivariate model containing all the variables that resulted from the previous step. Then, by looking at the p-value of the Wald statistic of each variable we should eliminate those that do not contribute enough. This will generate a smaller model that should then be compared to the previous one through the partial likelihood ratio test.

The elimination process mentioned in step 2 is not perfect. Hence, in step 3, we must compare the coefficients of each variable of the new smaller model to the coefficient on the previous large model. If there is a change greater than 20%, it means an important variable was removed in the previous step, and therefore it should be added back. We should repeat steps 2 and 3, removing and adding one or few variables at each time, until we reach a final stage where all the variables in the model are clinically or statistically important.

In step 4, we should add to the model that resulted from the previous steps each variable that was rejected in the first step, one at a time, checking the significance of each one through the Wald statistic p-value or the partial likelihood ratio test to verify if it is a categorical value with more than two levels. The reason for this is that some variables might not be significant by themselves, but when regressed together with others they may substantially contribute to and actually improve the model.

Step 5 requires us to have a closer look at each variable and test the assumption that the logit increases or decreases linearly as a function of the covariate. By the end of this process we have what it is called the main effects model (Hosmer, Lemeshow, & Sturdivant, 2013).

In step 6 we complement the model with the addition of variable transformations such as interactions. As discussed before, the decision to add or not a certain interaction has to be based on statistical as well as practical considerations. Although, in the business project we concluded that the addition of such transformations had a high potential, only now that we know the meaning of each variable we can see if it really improves the model or not. Interactions add the ability to measure the impact of a variable conditional to the level of another variable. Even if there is statistical evidence that the variable adds explanatory power, it should only be added to the model if it makes sense from a clinical perspective. According to the work of Hosmer et al. (2013, p.92), "we address the clinical plausibility issue by creating a list of possible pairs of variables in the model that have some realistic possibility of interacting with each other". In fact, we should only test statistically the interactions that make economic sense. We should perform all the previous steps for the interactions.

Finally, in step 7, we must test the fit of the final model and assess its adequacy through the traditional goodness of fit methods.[2]

2.3.3. Ideal number of variables

The ideal number of variables a model should have is an important discussion among the modeling experts. Some posit that, independently of their statistical significance, the more variables one adds to a model the better it is, as long as the variables make economic sense (Rothman, Greenland, & Lash, 2008). While other researchers defend that the model should be simplified as much as possible to a point where from the whole selection of variables, only those which have statistical significance should be kept. In fact, there are advantages and disadvantages of having more or less variables and for the purpose of risk management, a simple model could make more sense as it is more likely to be more stable – a fundamental requirement of the Basel Accord. Additionally, it becomes easier and more economical to implement as less data have to be collected, stored and monitored.

The rationale behind the previous statement is the following: as the number of variables increase, the estimated standard errors become greater, and consequently the model becomes more dependent on the observed data. Furthermore, adding too many variables in the model may produce numerical unstable estimates due to “overfitting”. This means that the estimated coefficients or the estimated standard errors may be unrealistically large. Finally, as the number of variables increase, the size of the sample needs to be larger, which is an important drawback as acquiring reliable data is a bottleneck for the UK MBS’s.

2.4. Conclusions

The most important step in building a logistic regression model to estimate the probability of default is the selection of the explanatory variables. As discussed above, this selection process should be divided into two parts: (a) a research of potential variables to be input in the model in terms of clinical importance that, in the case of MS, should comprise borrower and mortgage specific parameters as well as relevant macroeconomic variables; and (b) the statistical analysis of those inputs, to reach the final set of variables, following a structured process like the one presented in this paper. We can conclude that for the purpose of risk management, a small set of covariates is recommended for the sake of simplicity and stability.

[2] This subject was discussed in the Business Project

3. REFLECTION ON LEARNING

The business project is one of the cornerstones of the CEMS MIM Program. It is an important way to apply the knowledge acquired during the program to a real life situation. All in all, it is a tremendous learning experience and, in this section, I reflect on that.

3.1. Masters Content Applied

Many times, students question themselves why they have to study certain topics, which they believe don't add any value to their education. The answer to that question becomes very clear when one gets a closer look at a real company while working on a real project. A master's degree is not only about a few specific topics that we care about the most. It is a whole package that gives us a strong background to allow us to understand the world quickly. By completing a MSc. in Finance, we undoubtedly gain a solid foundation in finance; however the biggest asset that we acquire is the ability to absorb new information rapidly. And it is this ability to quickly assimilate knowledge and take conclusions to make decisions that makes us valuable in the job market. That was exactly what I felt during my project with MS. Although the BP was about risk management, a subject I never took at university, I could catch up very fast and have a clear idea of what it is that we were talking about.

Of course, the specific knowledge that I acquired in certain subjects I took were also very valuable. It would have been much harder to deliver significant value in this project without having taken Financial Econometrics. I not only applied most of the topics I learnt from this course, but also went further using the "additional learning" resources that we received there. Moreover, other courses like Investments, Macroeconomics of Financial Markets were naturally very helpful and provided me with knowledge I could directly apply.

Another very important lesson, regarding my master's degree, was the fact that even topics that were completely unrelated to risk management and even finance played an important role. For instance, during Customer Relationship Management (CRM) we worked with a large database of customers that we used to apply data mining techniques. During my BP, I found myself using some of those techniques to further explore the data provided by MS.

Another aspect in which the master's degree content definitely helped me was the constant development of soft skills throughout the program. Presentation skills, team work, negotiation, emotional intelligence are a few from the vast list of soft skills that the master's degree helped me to develop. These skills are game-changers and absolutely contributed to the success of my business project.

3.2. New Knowledge

No matter how much you study or how many courses you take, there will always be new things to learn. In this project I had the chance to take my statistical analysis skills to a whole new level by learning SPSS and STATA. I can now confidently say that I am an advanced user of both software solutions that I had never used before. These are tools that are likely to be very valuable for my professional future.

Additionally, I deepened my understanding about regression models and how they can be applied in real-life situations. Social Sciences courses focus mostly on the linear regression; in this project, however, I had the chance to learn other types of regressions in a detailed way, such as the logistic regression or the probit regression. And this is just one of vast set of the new concepts I learnt. The business project was a valuable learning source.

But the knowledge acquired in such a project goes beyond theoretical concepts or new tools. I learnt the reality of a real company. The fact that there is limited amount of resources and time and therefore choices must be made. The fact that I must influence people to get somewhere, or let go my ambition of doing something and pursue a better idea suggested by someone else. Working is only about relationships and my fortunes will be tied to others.

3.3. Personal Experience

When reflecting on the BP I can clearly identify some of my strengths and weaknesses. On the strengths side, I realized I had a greater capacity to work under pressure when compared to the rest of the team. Also I was very proactive. On the weaknesses, I feel I need to improve my time management skills by being more selective on how I send my time in projects. Being aware of my weaknesses is the most important step to improve, however. In future projects I will plan with more detail and do more check-up meetings to ensure I'm not wasting time.

3.4. Benefit of hindsight

When looking back, I realize there were two things I would have done differently. Firstly, we started working the data quite late. The idea was to get theoretical background in the first place to then apply it to the data all at once. I would rather start with the data, topic by topic, as the results of the practical application can totally change the direction of the research. Secondly, instead of starting to write a report, I would start by building the presentation. The reason is that it helps following a logical storyline, rather than just doing analysis by blocks, that guides the report to much more interesting conclusions. The results of such process are much more valuable.

4. REFERENCES

Basel Committee on Banking Supervision. (2006). *International Convergence of Capital Measurement and Capital Standards.* Basel, Switzerland: Bank for International Settlements · Press & Communications.

Blanco, A., Irimia, A., & Olive, M. D. (2012). The Prediction of Bankruptcy of Small Firms in the UK using Logistic Regression. *Análisis Financeiro*(118), pp. 32-40.

Campbell, J. Y., & Cocco, J. F. (2011, October). A Model of Mortgage Default. Harvard University & London Business School.

Donchev, T. (2009, August). Modeling Defaults in Residential Mortgage Backed Securities: An Intensity Based Approach. Amsterdam: Vrije Universiteit Amsterdam.

Engelmann, B., & Porath, D. (2012, February 16). Do not Forget the Economy when Estimating Default Probabilities. *Wilmott magazine*, pp. 71-73.

Falkenstein, E. (2000). *RiskCakc for Private Companies: Moody's Default Model.* New York: Moody’s Investors Service, Inc.

Fang, K., & Huang, H. (2011). Variable Selection for Credit Risk Model Using Data Mining Technique. *Journal of Computers*, VOL. 6, NO. 9.

Hosmer, D. W., Lemeshow, S., & Sturdivant, R. X. (2013). *Applied Logistic Regression.* Hoboken, New Jersey: John Wiley & Sons, Inc.

Neuenschwander, C., & Proffitt, D. (2013, January). Loan Characteristics, Borrower Characteristics, or Housing Price Declines: What Drives Mortgage Defaults? Mount Vernon Nazarene University & Arizona Christian University.

Rothman, J., Greenland, S., & Lash, T. (2008). *Modern Epidemiology.* Philadelphia: Wolters Kluwer/Lippincott Williams & Wilkins.

The Economist. (2014, January 11). It's Back. *The Economist*.

The Financial Times. (2014, March 27). *League Tables - Investment Banking Review*. Retrieved May 22, 2014, from The Financial Times: http://markets.ft.com/investmentBanking/tablesAndTrends.asp?ftauth=1400758737866